# Copula-Based Spatial Modelling of Geometallurgical Variables

G N Musafer[1,5], M H Thompson[2,5], E Kozan[3,5] and R C Wolff[4,5]


## ABSTRACT

The most important aspect of modelling a geological variable, such as metal grade, is the spatial correlation. Spatial correlation describes the relationship between realisations of a geological variable sampled at different locations. Any method for spatially modelling such a variable should be capable of accurately estimating the true spatial correlation. Conventional kriged models are the most commonly used in mining for estimating grade or other variables at unsampled locations, and these models use the variogram or covariance function to model the spatial correlations in the process of estimation. However, this usage assumes the relationships of the observations of the variable of interest at nearby locations are only influenced by the vector distance between the locations. This means that these models assume linear spatial correlation of grade. In reality, the relationship with an observation of grade at a nearby location may be influenced by both distance between the locations and the *value* of the observations (ie non-linear spatial correlation, such as may exist for variables of interest in geometallurgy). Hence this may lead to inaccurate estimation of the ore reserve if a kriged model is used for estimating grade of unsampled locations when non-linear spatial correlation is present. Copula-based methods, which are widely used in financial and actuarial modelling to quantify the non-linear dependence structures, may offer a solution. This method was introduced by Bárdossy and Li (2008) to geostatistical modelling to quantify the non-linear spatial dependence structure in a groundwater quality measurement network. Their copula-based spatial modelling is applied in this research paper to estimate the grade of 3D blocks. Furthermore, real-world mining data is used to validate this model. These copula-based grade estimates are compared with the results of conventional ordinary and lognormal kriging to present the reliability of this method.


## INTRODUCTION

In the evolution of geometallurgy, a key issue is the estimation of geometallurgical variables. There has been some debate as to the estimation of variables which have non-linear dependence and more complex spatial structures than are generally allowed for in conventional linear geostatistics. This contribution presents some early attempts to provide alternative spatial estimators, based on copulas, which may prove fertile for the development of the required estimation approaches. While the example followed here is a metal grade, the concept could have important applications for geometallurgical variables, such as mineralogy and hardness.

The most important aspect of dealing with a geological variable, such as metal grade, is the spatial correlation. Spatial correlation describes the relationship between realisations of a geological variable sampled at different locations. Any method modelling a geological variable should be capable of accurately estimating the true spatial correlation.

The variogram and covariance function are the most common methods used to capture the spatial dependence structure of a geological variable (Gräler and Pebesma, 2011; Kazianka and Pilz, 2010a). These methods assume linear dependence at any lag distance. In other words, these methods assume that spatial dependence only depends on distance between the locations of the observations. Moreover, kriged models, which are most commonly used to model geological variables, extract the spatial dependence through the variogram. However, in reality, in most cases the spatial dependence structure may vary over the different percentile values of the distribution of the variable of interest (Journel and Alabert, 1989). Consequently, the spatial dependence structure may be influenced by other factors, such as the value of observations (Journel and Alabert, 1989). Therefore the kriged model may produce inaccurate estimators of distributional properties of the variable at unsampled locations when a non-linear dependence structure is present. Bárdossy and Li (2008) introduced a new geostatistical model based on copulas. Copulas are able to capture any non-linear dependence structure over the distribution of the variable of interest. Hence copulas consider both the spatial geometry


1. PhD Candidate, School of Mathematical Sciences, Queensland University of Technology, 2 George Street, Brisbane Qld 4000. Email: g.musafer@qut.edu.au
2. Lecturer in Statistics, School of Mathematical Sciences, Queensland University of Technology, 2 George Street, Brisbane Qld 4000. Email: helen.thompson@qut.edu.au
3. Professor and Head of Decision Sciences, School of Mathematical Sciences, Queensland University of Technology, 2 George Street, Brisbane Qld 4000. Email: e.kozan@qut.edu.au
4. Professor and Chair in Geostatistics and Mine Planning, WH Bryan Mining and Geology Research Centre, The University of Queensland, Brisbane Qld 4072. Email: rodney.wolff@uq.edu.au
5. Cooperative Research Centre for Optimising Resource Extraction (CRC ORE)

and the value of the observation when modelling spatial dependence structure (Haslauer, Li and Bárdossy, 2010).

As far as the authors are aware, copula models have been used in only a few spatial applications, for example to model hydrology properties (Bárdossy and Li, 2008), soil properties (Marchant *et al*, 2011) and air pollutants (Kazianka and Pilz, 2011). Even though copula-based modelling is a new avenue for spatial statistics, it has been widely used in non-spatial applications in fields where it is essential to deal with non-linear dependence, such as in finance and actuarial sciences (Bárdossy, 2006).

The main objectives of our research are to fit copula-based models in order to estimate the grade of a metal obtained from a real mine site and to estimate the distribution of the grade of metal at unsampled locations, conditional on the local neighbourhood of unsampled locations. Moreover, the Gaussian copula-based model and the Student t copula-based model are compared with a kriged model and lognormal kriged model to evaluate the reliability of the copula-based models.

This paper contains four sections. The copula section describes the basic properties of copulas; the following section presents the copula-based geostatistical modelling. Within the application section copula-based modelling is applied to large-scale mine data and the results of the data are given within the same section. The last section is devoted to conclusions driven by the results and discussion of copula-based modelling.

## COPULAS

An introduction to copula theory can be found in Nelsen (2007) and Trivedi and Zimmer (2007). For an applied review of copulas, the reader is referred to Boardmana and Vann (2011).

A copula can be defined as a multivariate distribution function of uniformly distributed random variables on the interval [0,1]. Therefore it has the same properties as any distribution function. For a multivariate distribution function $C(u_1,\ldots,u_n)$ to be a copula, it should satisfy the following conditions:

$$U_1,\ldots,U_n \sim Uniform(0,1)$$

$$C(1,\ldots 1,u_i,1,\ldots 1) = u_i \text{ for every } i \leq n \text{ in } [0,1]$$

$$C(u_1,\ldots, u_n) = 0 \text{ if } u_i = 0 \text{ for any } i \leq n$$

$C$ is an $n$-increasing function

*Sklar's theorem* describes the relationship between the copula and the joint distribution function $F(z_1,\ldots,z_n)$ of an $n$-dimensional vector of random variables $(z_1,\ldots,z_n)$ as follows:

$$F(z_1,\ldots,z_n) = C(F_1(z_1),\ldots,F_n(z_n))$$

where:

$F_i(z_i)$ represents the $i^{th}$ one-dimensional marginal distribution function

This theorem states that, for a given joint distribution, a copula can be found to model the multivariate structure of a vector of random variables by using their marginal distributions. Moreover, the copula will be unique if all the marginal distributions are continuous.

Based on Sklar's theorem, the joint density function $f(z_1,\ldots,z_n)$ of the random variables $(Z_1,\ldots,Z_n)$ can be derived by applying partial derivation on joint density function $F(z_1,\ldots,z_n)$. Hence, the joint density function can be fragmented into its univariate margins and dependence structure as follows:

$$f(z_1, \ldots, z_n) = c(u_1, \ldots, u_n \times \prod_{i=1}^{n} f_i(z_i)$$

where:

$u_i$ is $F_i(z_i)$ for $i = 1,\ldots,n$

$c(u_1,\ldots,u_n)$ denotes the density function of the copula

In other words, the copula density can be expressed as the dependence structure of the random vector $Z_1,\ldots,Z_n$.

## COPULA-BASED GEOSTATISTICAL MODELLING

The definitions of terminology used in this section can be found in the appendix.

### Assumptions

As with conventional geostatistics, copula-based modelling assumes that the set of measured values of the variable of interest are realisations of a random function (Bárdossy and Li, 2008). However, when applying copula-based models, a strong stationary random function is assumed over the domain of interest. This assumption is stronger than the conventional linear geostatistical assumption of a second order stationary random function over the domain of interest. However, copula-based modelling has more advantages when compared to geostatistical modelling, even though it requires a more limiting assumption, such as the capability to obtain the full conditional distribution, ability to remove the influences of marginal distributions when modelling the dependence structure and ability to model the non-linear spatial dependence (Haslauer, Li and Bárdossy, 2010). Moreover, this stationary assumption is highly influenced by the geological nature such as different rock types of the domain. In some situations it is required to subdivide the domain of interest into statistically homogeneous subdomains with common geological properties (Larrondo and Deutsch, 2005). By applying the model to these subdomains separately, stationary assumptions, which are required for the model, can be secured. Thus in the application of copula-based models, the geological nature of the domain must be deeply analysed and as mentioned above, the subdivision process should be applied if required.

Based on this strong stationary assumption, the marginal distributions of the variable of interest for each location in the domain are identical, ie $F_i(z_i) = F(z_i)$. The empirical bivariate copula can be used to explore the spatial variability. As with the variogram, it is assumed that the bivariate spatial copula $C_s$ at any two locations only depends on the separation vector $\vec{h}$ (and is independent of the locations $x$) (Bárdossy, 2006; Bárdossy and Li, 2008); that is:

$$\begin{aligned} C_s(u,v) &= Pr(F(Z(x)) < u, F(Z(x+\vec{h})) < v) \\ &= C_{\vec{h}}(F(Z(x)), F(Z(x+\vec{h}))) \end{aligned}$$

Moreover, not just any copula model can be used as a spatial copula (Bárdossy and Li, 2008). There are requirements that should be fulfilled by the copulas to be a spatial copula

(Bárdossy and Li, 2008; Kazianka and Pilz, 2010b). Generally, as with conventional geostatistics, it is assumed that the spatial dependence between location $x_1$ and location $x_2$ is the same as the dependence between location $x_2$ and location $x_1$. Hence, this symmetrical property should be able to be also described by the spatial copula. Another requirement is that the dependence structure of the copula must be able to be parameterised in order to be described as a function of the vector $\vec{h}$. Furthermore, the well-known spatial property of no dependence between far distant observations and high dependence between near observations property can be represented in copula-based models as follows:

$$C(u_1, \ldots, u_n) = \prod_{i=1}^{n} u_i \text{ when } \|h\| \to \infty$$
$$C(u_1, \ldots, u_n) = \min(u_1, \ldots, u_n) \text{ when } \|h\| \to 0$$

## Copula-based spatial analysis

Based on the papers of Bárdossy *et al* (2008) and Kazianka and Pilz (2010a), the procedure for spatial analysis using copulas is described as follows. Even though this model can be applied to anisotropic cases, in this application we only consider the isotropic case. In isotropic situations, it is assumed that spatial dependence varies only with the distance not on the direction. In this case the vector $\vec{h}$ becomes distance $h$.

### *Step 1: explore the spatial dependence structure*

The first step in spatial analysis is construction of a bivariate copula to explore the dependence structure.

As mentioned above, the marginal univariate distributions of the variable of interest for each location are identical (based on the stationary assumption). Therefore, the empirical marginal distribution function $F(z)$ can be estimated using all the observations $z(x_1),\ldots,z(x_N)$. Then the set of pairs $S(h)=\{F(z(x_i)), F(z(x_j))\}$ where $x_i - x_j \simeq h$ or $x_j - x_i \simeq h$ can be calculated. All the points in $S(h)$ are in the unit square. As a result of the symmetry property of the dependence structure, if any $(u_1, u_2) \in S(h)$ then $(u_2, u_1) \in S(h)$.

The empirical bivariate copula densities can be calculated using a kernel density smoothing method if the number of pairs is large enough; otherwise, the empirical bivariate copula can be calculated by defining a regular grid on the unit square and calculating the cumulative frequency of values for each grid.

### *Step 2: parameter estimation*

There are three types of parameters that should be estimated in the spatial copula model: correlation structure $\theta$, copula parameters $\lambda$ and the parameters of the marginal distribution $\alpha$. These parameters can be estimated using the method of maximum likelihood (Kazianka and Pilz, 2010b).

The likelihood function is given by:

$$L(\Theta; Z(x)) = c_{\theta,\lambda}(F_\alpha(Z_1), \ldots, F_\alpha(Z_N)) \times \prod_{i=1}^{n} f_\alpha(Z_i)$$

where:

$\Theta = (\theta,\lambda,\alpha)$

$c\theta,\lambda$ is the copula density

$f_\alpha$ is the marginal density

$F_\alpha$ is its distribution function

If the copula is Gaussian or Student t, then there are no difficulties in applying the maximum likelihood method directly. However, calculation of the copula density for higher dimensions may be difficult if the copula is a non-central chi-squared copula (which was introduced by Bárdossy (2006)). As a solution to this, Kazianka and Pilz (2010a) propose using the bivariate copula densities to perform the maximum likelihood estimation based on the assumption of independence of different pairs of observations.

However, Gaussian and Student t copula models are only used in this paper due to the demanding computational requirements when fitting non-central chi-squared copulas to large scale data sets. For example, if $n$ is the number of observations, $2^n$ terms of calculations are needed in the process of spatial interpolation to estimate the value of the variable of interest at unsampled location.

### *Step 3: spatial interpolation*

The final aim of any spatial analysis method is to estimate the variable of interest $Z(x)$ at an unsampled location $x$. Although the kriging estimator is able to produce the expected value at the unsampled location as the estimator, the methodology introduced by Bárdossy and Li (2008) allows one to estimate the full conditional distribution of the $Z(x)$, which is:

$$F(Z(x) \mid Z(x_1) = z(x_1),\ldots,Z(x_N) = z_N) = Pr(Z(x)<z \mid Z(x_1) = z_1,\ldots,Z(x_N) = z_N)$$

where:

$N$ is the total number of observations

The full conditional distribution of the variable of interest at an unsampled location can be written using the corresponding conditional copula $C_{x,N}$:

$$F(Z(x) \mid Z(x_1) = z(x_1),\ldots,Z(x_N) = z_N) = C_{x,N}(F(Z(x)) \mid u_1 = F(Z(x_1)),\ldots,u_N = F(Z(x_N)))$$

However, it may be computationally difficult to use all the observations in this process. Therefore, the conditional distribution is obtained based on the local nearby points (Bárdossy and Li, 2008).

Let $n$ be the nearby points to the unsampled location $x$. Then:

$$F(Z(x) \mid Z(x_1) = z(x_1),\ldots,Z(x_n) = z_n) = C_{x,n}(F_\alpha(Z(x)) \mid u_1 = F_\alpha(Z(x_1)),\ldots,u_n = F_\alpha(Z(x_n)))$$

Therefore the conditional density function can be derived as:

$$f(z \mid z_1, \ldots, z_n, \Theta) = \frac{\partial F(Z(x) \mid Z(x_1) = z_1, \ldots, Z(x_n) = z_n)}{\partial z}$$
$$\frac{\partial C(u \mid u_1 = F(Z(x_1)), \ldots, u_n = F(Z(x_1)))}{\partial z}$$
$$\frac{\partial C(u \mid u_1 = F(Z(x_1)), \ldots, u_n = F(Z(x_1)))}{\partial u} \times \frac{\partial F(z)}{\partial z}$$

that is:

$$f(z|z_1,\ldots,z_n,\Theta) = c(u|u_1 = F(Z(x_1)),\ldots,u_n = F(Z(x_1)))\times f(z) \quad (1)$$

It is clear that in order to construct the conditional density function of the variable of interest at an unsampled location, constructing a conditional copula density is essential. The procedure for constructing the copula density is described as follows:

1. Transform the observed $z_i$ to $z_i'$ by using the inverse univariate distribution function $H^{-1}_{z'}(F_\alpha(z_i))$ of the multivariate joint distribution used to generate the prospective copula. For an example, if the prospective copula is Gaussian then $H^{-1}_{z'}$ is the standard normal distribution function.
2. To estimate the value of the variable of interest at unsampled location, a total of $n$ nearby sampled locations are selected and their transformed values are used in the constructed conditional copula density:

$$c_n(u \mid u_1 = H_{z'}(z'_1), \ldots, u_n = H_{z'}(z'_n)) = \frac{c(u = H_{z'}(z'), u_1 = H_{z'}(z'_1), \ldots, u_n = H_{z'}(z'_n))}{c(u_1 = H_{z'}(z'_1), \ldots, u_n = H_{z'}(z'_n))}$$

Since $c(u_1 = H_{z'}(z'_1), \ldots, u_n = H_{z'}(z'_n))$ is a constant:

$$c_n(u \mid u_1 = H_{z'}(z'_1), \ldots, u_n = H_{z'}(z'_n)) \propto c(u = H_{z'}(z'), H_{z'}(z'_1), \ldots, u_n = H_{z'}(z'_n)) = \frac{h_{n+1}(z', z'_1, \ldots, z'_n)}{h_1(z') * h_1(z'_1) * \ldots * h_1(z'_n)}$$

where:

$h_{n+1}$ is the $n+1$ dimensional joint density of the $Z'$

Since transformation $H^{-1}_{z'}(F_\alpha(z_i))$ is rank preserving:

$$c_n(u \mid u_1 = F(Z(x_1)), \ldots, u_n = F(Z(x_n))) = c(u \mid u_1 = H_{z'}(z'_1), \ldots, u_n = H_{z'}(z'_n))$$

Therefore, Equation 1 can be written as:

$$f(z \mid z_1, \ldots, z_n, \Theta) = c_n(u \mid u_1 = H_{z'}(z'_1), \ldots, u_n = H_{z'}(z'_n)) \times f_\alpha(z) \tag{2}$$

Now any estimator can be calculated for the unsampled location using the conditional distribution.

As an example, the expected value or the median of the conditional distribution can be obtained using Equations 3 and 4, respectively:

$$\text{Expected value} = \int_{-\infty}^{\infty} z \times c(u \mid u_1 = H_{z'}(z'_1), \ldots, u_n) = H_{z'}(z'_n)) \times f_\alpha(z)\, dz = \int_0^1 F_\alpha^{-1}(u)\, c(u \mid u_1 = H_{z'}(z'_1), \ldots, u_n = H_{z'}(z'_n))\, du \tag{3}$$

$$\text{Median value} = F_\alpha^{-1}(u = C_n^{-1}(0.5)) \tag{4}$$

Since this method provides the full conditional distribution at an unsampled location, it is easy to obtain a more complete estimation of uncertainty, such as confidence intervals, when compared to the kriged model. Here 'complete' is used to emphasise that the copula-based model is fully capable of producing uncertainty estimation dependent on both the observations' configuration and values (Bárdossy and Li, 2008; Haslauer, Li and Bárdossy, 2010). This feature is very important for additional drilling campaigns, where a reduction of uncertainty is expected based on the influence of additional measurements.

## APPLICATION

In this section copula-based modelling is applied to a real-world mine data set. There are nearly 80 000 measurements in the data set from over 2000 drill holes. A small-scale example is presented here based on a subset of the spatial observations and spatial analysis performed on this subset for the grade of the main metal. The random subset should retain the statistical properties important to the study, ie the distributional properties of the main metal. Our validation steps (below) suggest this assumption is reasonable. Consequently, our demonstration of copula modelling and comparison to kriged methods should also be applicable to the full data set of measurements.

This selected subset consists of 13 832 measurements of grade of the main metal $z(x_i)$ at three dimensional locations $x_i = (x_{1i}, x_{2i}, x_{3i}), i = 1,\ldots,13\,832$. For confidentiality reasons we are unable to present the spatial plot of the metal grades, but we offer some summary statistics of the grade of the main metal for both the exhaustive data set and selected subset in Table 1. In Table 1 it can be seen that the mean and median of the subset compares well with the exhaustive mean and median respectively. The coefficient of variation (CV) of the subset is very similar to the CV of the exhaustive data. Even though skewness differs, the skewness coefficient is a somewhat non-robust measure and the skewness for both data sets is positive. Figure 1 displays the histograms of the metal grade for the exhaustive data set and the subset respectively, from which the skewness is clearly demonstrated. Finally, the comparisons of the histograms and the summary statistics in Table 1 displays evidence that the selected subset can be considered to be a representative subset of the exhaustive data set.

**TABLE 1**

Summary statistics of the grade of the main metal for both the exhaustive data set and selected subset.

| Method | Exhaustive | Subset |
|---|---|---|
| n | 78 256 | 13 832 |
| Mean | 0.7483061 | 0.8544585 |
| Standard deviation | 1.114259 | 1.200226 |
| Coefficient of variation | 1.489042 | 1.404663 |
| Skewness | 2.26383 | 1.840481 |
| Min | 0.005 | 0.005 |
| First quartile (Q1) | 0.03 | 0.025 |
| Median | 0.186 | 0.15 |
| Third quartile (Q3) | 1.092 | 1.4 |
| Max | 11.148 | 8.863 |

The first step of the copula-based spatial analysis is estimating the marginal distribution function of the variable of interest, $Z(x)$. By using all of the observations of metal grades in the selected subset, the marginal distribution was estimated based on the maximum likelihood value. The model which produces the largest maximised likelihood is selected as the best fit. The log-likelihood values for the Gaussian, generalised extreme value distribution, lognormal distribution and gamma distribution were -22151.0, -6062.1, -6032.5 and -6352.3, respectively. Therefore, the lognormal distribution was selected as the best fitting distribution among the competing models. The estimated location parameter for the fitted distribution is 1.7342 and the estimated scale parameter is 2.1201.

As mentioned above, it is difficult to use all locations in the estimation process. Therefore a sufficient number of nearby

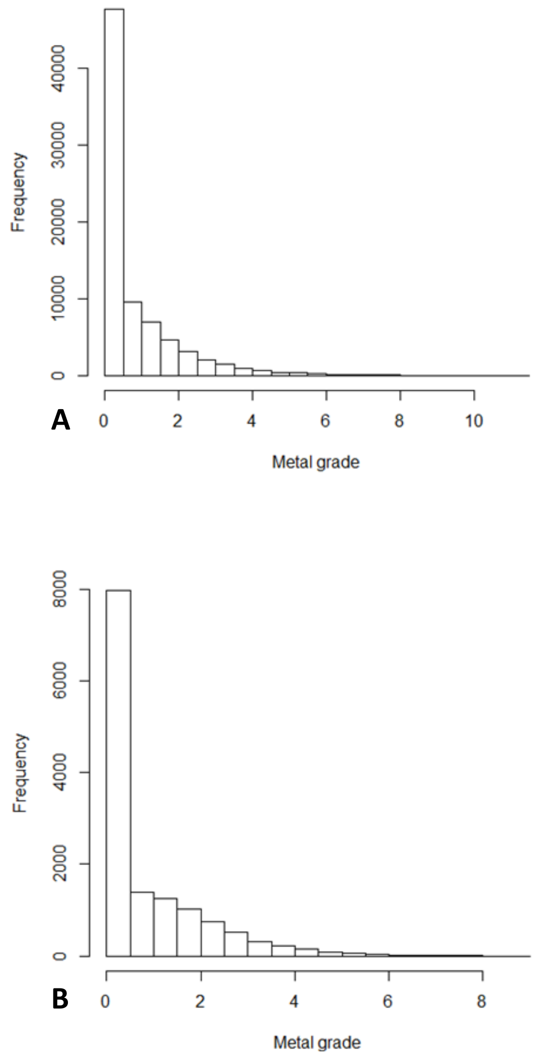


**FIG 1** - (A) Histogram of metal grade of exhaustive data set; (B) histogram of metal grade of subset.

locations has to be selected to produce good estimators. Bárdossy and Li (2008) explain the method of selecting a sufficient number of nearby locations. For a few randomly selected locations, the density functions were estimated and plotted for different numbers of nearby locations (ie 3, 6, 8, 10, 13, 17 and 20). Nearly identical density functions are observed after eight nearby locations for almost all considered locations. Therefore any number of points greater than eight can be considered as a sufficient number of nearby locations for an example. Hence, we use ten nearby locations are in our estimation process.

For simplicity, the Gaussian copula and Student t copula with two degrees of freedom were used as competing copula models, and the spherical correlation function was used to model the spatial correlation based on the maximum likelihood method. Gaussian, exponential correlation and Marten (Pardo-Iguzquiza and Chica-Olmo, 2008) models were the other models which competed with the spherical model. The estimated nugget value of the spherical model is 0.1204 and the estimated range is 37.0029.

The four spatial models that were compared are:

1. ordinary kriged
2. lognormal kriged
3. Gaussian copula-based model
4. Student t copula model.

An isotropic random field was assumed for all the models investigated in this paper and the models' performances are compared based on the point estimations.

## Cross-validation

Cross-validation was carried out to compare the performance of the four selected spatial models. The leave-one-out cross-validation technique was used, with ten nearby locations used for each location when constructing the conditional copula in the interpolation process. Two estimators, mean and median, were used for the copula models. The performances of the models were evaluated using different criteria: mean absolute error (MAE) and per cent bias (bias percentage between estimated and true value).

Figure 2 shows the scatter plot of the true values and the estimated values by cross-validation using the above-mentioned four models. According to Table 2, even though median estimation of the Student t copula-based model has the lowest MAE when compared to other estimators, it has unacceptable bias percentage from a practical viewpoint. In terms of MAE, the second best estimator is the median estimation of the Gaussian copula-based model. Even though the MAE of the ordinary kriged model is slightly higher than the median estimator of the Gaussian copula, the ordinary kriged model has the smallest negligible bias percentage. Additionally, it can be seen that, even though the data followed the lognormal distribution, lognormal kriging has the highest MAE. This is the kind of cost which has to be paid for the 'back transformation' of lognormal kriging (Dowd, 1982; Roth, 1998). If lognormal kriging has overestimated the standard error at unsampled location, when it back transforms to the original scale, it becomes larger due to exponentiation and finally it will lead to a seriously overestimated prediction (Roth, 1998).

**TABLE 2**

Cross-validation result for the interpolation of different spatial models.

| Method | Mean absolute error | Per cent bias (%) |
|---|---|---|
| Gaussian copula – mean | 0.4164148 | 2.24 |
| Gaussian copula – median | 0.3572973 | -21.25 |
| t copula – mean | 0.4256659 | 17.67 |
| t copula – median | 0.3453207 | -15.73 |
| Kriging | 0.3578423 | 0.47 |
| Log kriging | 0.4672146 | 4.36 |

## DISCUSSION AND CONCLUSION

Even though, in this paper, only two simple copula-based models have been used, the median estimator of Student t copula model and the median estimator of Gaussian copula model were able to perform better than the kriged model in terms of MAE. However, in mining application the mean estimator is expected to perform well because it has the ability to produce an unbiased estimator for total metal content. Below we suggest solutions that may lead to the improvements of a mean estimator for the copula-based model.

Due to computational difficulties, non-central chi copulas have not been applied. Although the non-central chi-squared method is better for modelling asymmetric dependence, it is unable to capture tail dependence. As with asymmetric dependence, tail dependence structures occur frequently

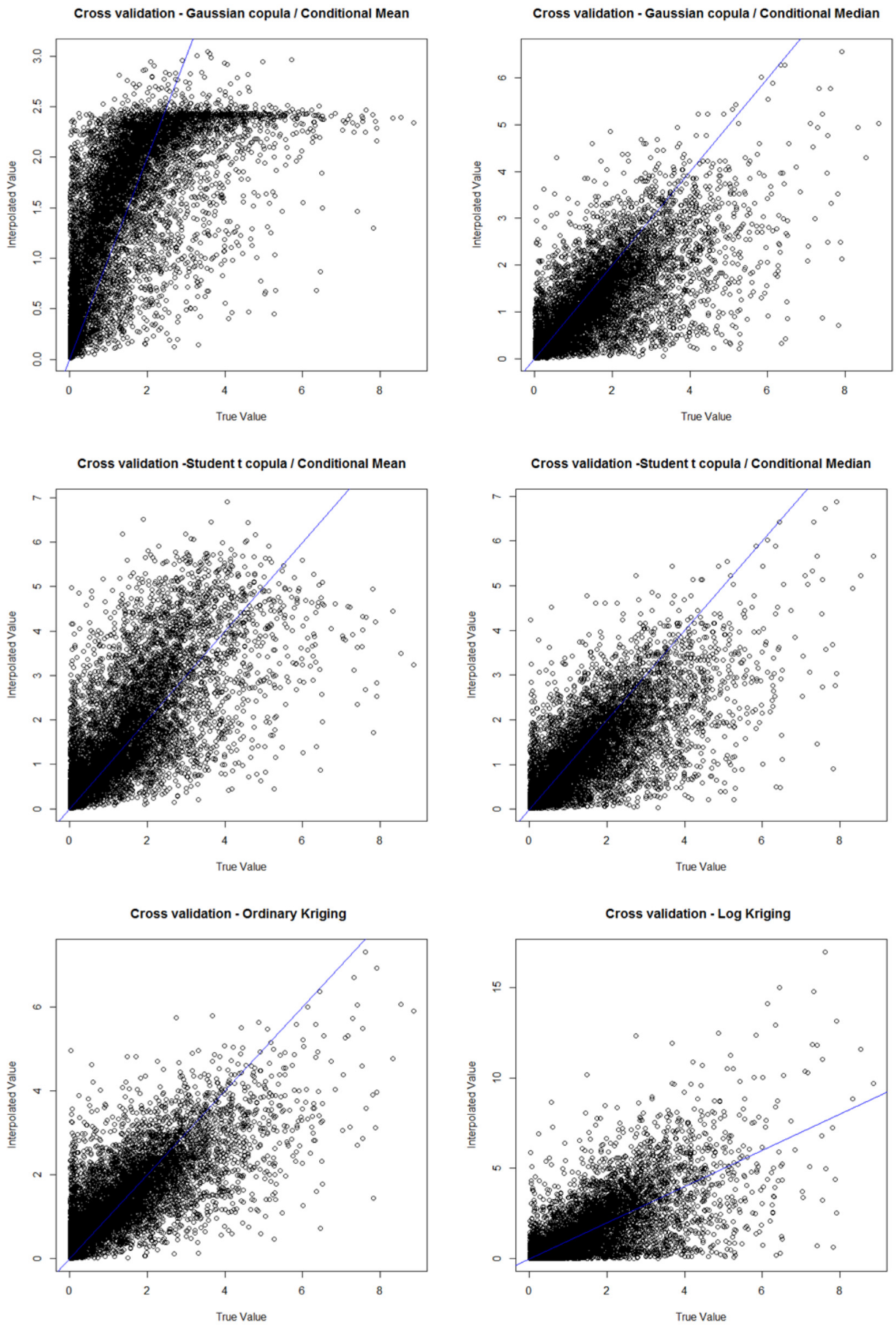


**FIG 2** - Scatter plot for cross-validation result for interpolation of different models.

in spatial data. However, the Gaussian copula model is not capable of modelling either asymmetric dependence structures or tail dependence. However, Student t copulas are able to capture the tail dependence structure if the reflection symmetry assumption is satisfied. This may be the reason for the good performance of the Student t copula even if it lacks the capability to extract asymmetrical dependence structure. According to Figure 3, the empirical bivariate copula shows a high concentration of the probability at the tails of the distribution of the metal grade for the different lag distances. This is an indication of the existence of a tail dependence structure for the metal grade of metal of interest in this research.

Copulas have potential to become a popular geostatistical model because of the capability to fit the full conditional distribution, the ability to remove the influences of marginal distributions when modelling the dependence structure and the ability to model non-linear spatial dependence. Even though this research paper focuses on modelling grade, this method can be used to model any geoscientific variable. Moreover, throughout this process we considered estimation of the value for the variable interest based on point support.

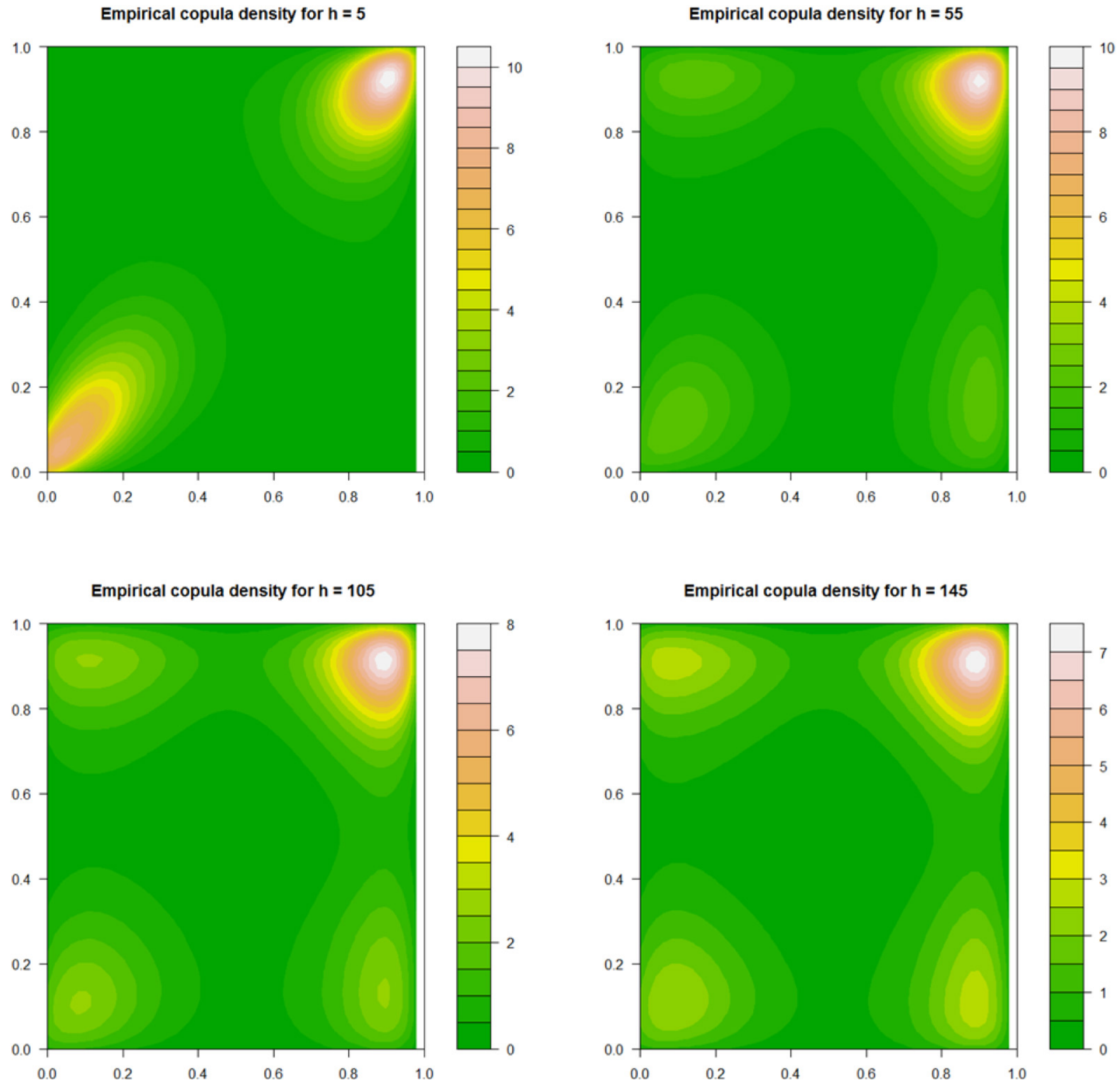


**FIG 3** - Bivariate copula densities for lag distance classes 5, 55, 105 and 145.

However, in mining, it is essential to estimate the value of the variable interest based on block support. As with the kriged model, the estimate of the middle point of the block volume (which is equal to the average of the estimates of a large number of discretised points in the block volume) can be considered as the estimator of variable of interest for block support in the copula-based model. However the copula-based model is more informative than the kriged model. In other words, it is capable of estimating the conditional density of point support. Thus there may be a path to find the conditional density function of block support based on the conditional density function of point support. However, further research should be carried out to find out the optimal method of estimating the variable of interest based on the block support by using the copula-based model.

There is still only a small number of candidate models suitable for copula-based spatial modelling. Almost all of these families lack the ability to deal with tail spatial dependence, which is an important limitation. Another drawback of using existing copula models is that the same copula is assumed at each lag distance. For instance, there may be a situation where a bivariate copula may suggest that the Clayton copula is suitable for the first two lag classes and a non-central copula is suitable for the remaining lag classes. But using the existing simple copula procedure, this information cannot be included in the copula model. Graler and Pebesma (2011) have suggested a pair-copula construction procedure as a solution for this. We are currently adapting this method to construct a more flexible copula-based model which will be able to capture the *in situ* dependence structure of the variable of interest.

Finally, from the results, it can be concluded that even though the Student t copula-based model is not able to capture the asymmetric dependence, its median estimator produces comparable results to the kriged model in terms of MAE. On the other hand, the Student t copula-based model produces larger bias when compared to the kriged model. Altogether, the outcome from the Student t copula-based model suggests potentially significant improved performance of copula-based models over the kriged model if we can address some of the limitations discussed here in our future work; for example, the use of a pair-copula to capture the *in situ* dependence structure and an extended approach to deal with block support.

## ACKNOWLEDGEMENTS

The authors would like to acknowledge the support of CRC ORE, established and supported by the Australian Government's Cooperative Research Centres Programme. The authors also thank BHP Billiton and the Quantitative Group for generous participation in this research.

Moreover, computational (and/or data visualisation) resources and services used in this work were provided by the HPC and Research Support Group, Queensland University of Technology, Brisbane, Australia.

## APPENDIX

### Definition: strong stationary random function

Let $Z(x)$ be a random function and $P(Z(x_1)\leq z_1,\ldots,Z(x_n)\leq z_n$ be the cumulative distribution function of the joint distribution of $Z(x)$ at the locations $x_1,\ldots,x_n \in D$. $Z(x)$ is said to be a strong stationary random field if, for all $n$ and all vectors $h$ that satisfy $x_1+h,\ldots,x_n+h \in D$.

$$P(Z(x_1)\leq z_1,\ldots,Z(x_n)\leq z_n) = P(Z(x_1+h)\leq z_1,\ldots,Z(x_n+h)\leq z_n)$$

This implies that the cumulative distribution function is not a function of $h$.

### Definition: second order stationary random function

Let $Z(x)$ be a random function. $Z(x)$ is said to be a second order stationary random field if:

- it has a constant mean over all spatial locations ie $E[Z(x)] = \mu$
- the auto-covariance of the data generating process depends only on distance $h$; ie $Cov(Z(x+h),Z(x)) = \gamma(h)$, where $\gamma$ is the covariance function of $Z(x)$.

## REFERENCES

**Bárdossy,** A, 2006. Copula-based geostatistical models for groundwater quality parameters, *Water Resources Research*, 42:W11416.

**Bárdossy,** A and Li, J, 2008. Geostatistical interpolation using copulas, *Water Resources Research*, 44:W07412.

**Boardmana,** R and Vann, J, 2011. A review of the application of copulas to improve modelling of non-bigaussian bivariate relationships (with an example using geological data), in *Proceedings 19th International Congress on Modelling and Simulation* MODSIM2011, pp 627-633 (Perth).

**Dowd,** P, 1982. Lognormal kriging – the general case, *Journal of the International Association for Mathematical Geology*, 14:475-499.

**Gräler,** B and Pebesma, E, 2011. The pair-copula construction for spatial data: a new approach to model spatial dependency, *Procedia Environmental Sciences*, 7:206-211.

**Haslauer,** C P, Li, J and Bárdossy, A, 2010. Application of copulas in geostatistics, *geoENV VII–Geostatistics for Environmental Applications*, pp 395-404.

**Journel,** A G and Alabert, F, 1989. Non-Gaussian data expansion in the Earth Sciences, *Terra Nova*, 1:123-134.

**Kazianka,** H and Pilz, J, 2010a. Spatial interpolation using copula-based geostatistical models, *geoENV VII–Geostatistics for Environmental Applications*, pp 307-319.

**Kazianka,** H and Pilz, J, 2010b. Copula-based geostatistical modeling of continuous and discrete data including covariates, *Stochastic Environmental Research and Risk Assessment*, 24:661-673.

**Kazianka,** H and Pilz, J, 2011. Bayesian spatial modeling and interpolation using copulas, *Computers & Geosciences*, 37:310-319.

**Larrondo,** P and Deutsch, C V, 2005. *Accounting for Geological Boundaries in Geostatistical Modeling of Multiple Rock Types* (Springer: Netherlands).

**Marchant,** B, Saby, N, Jolivet, C, Arrouays, D and Lark, R, 2011. Spatial prediction of soil properties with copulas, *Geoderma*, 162:327-334.

**Nelsen,** R, 2007. *Springer Series in Statistics: An Introduction to Copula* (Springer US: New York).

**Pardo-Iguzquiza,** E and Chica-Olmo, M, 2008. Geostatistics with the Matern semivariogram model: a library of computer programs for inference, kriging and simulation, *Computers & Geosciences*, 34:1073-1079.

**Roth,** C, 1998. Is lognormal kriging suitable for local estimation?, *Mathematical Geology*, 30:999-1009.

**Trivedi,** P K and Zimmer, D M, 2007. *Copula Modeling: An Introduction for Practitioners* (Now Publishers Inc).